Phys Med Biol. 2014 Feb 7;59(4):1073-1095. [Epub ahead of print]

Transmission-less attenuation correction in time-of-flight PET: analysis of a discrete iterative algorithm.

Michel Defrise[1], Ahmadreza Rezaei[2], Johan Nuyts[3]

Abstract

The Maximum Likelihood Attenuation Correction Factors (MLACF) algorithm has been developed to calculate the maximum-likelihood estimate of the activity image and the attenuation sinogram in time-of-flight PET, using only emission data without prior information on the attenuation. We consider the case of a Poisson model of the data, in the absence of scatter or random background. In this case the maximization with respect to the attenuation factors can be achieved in a closed form and the MLACF algorithm works by updating the activity. Despite promising numerical results, the convergence of this algorithm has not been analysed. In this paper we derive the algorithm and demonstrate that the MLACF algorithm monotonically increases the likelihood, is asymptotically regular, and that the limit points of the iteration are stationary points of the likelihood. Because the problem is not convex, however, the limit points might be saddle points or local maxima. To obtain some empirical insight into the latter question, we present data obtained by applying MLACF to 2D simulated TOF data, using a large number of iterations and different initializations.

[1]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, B-1090, Brussels, Belgium,
mail: mdefrise@vub.ac.be,
[2]Dept. of Nuclear Medicine, Katholieke Universiteit Leuven, B-3000, Leuven, Belgium,
mail: ahmadreza.rezaei@uzleuven.be,
[3]Dept. of Nuclear Medicine, Katholieke Universiteit Leuven, B-3000, Leuven, Belgium,
mail: johan.nuyts@uzleuven.be.

1. Introduction

Positron emission tomography (PET) aims at estimating the spatial distribution of a tracer molecule labelled with a positron emitting isotope. This distribution, called the *activity image*, is reconstructed from the *emission data*, which consist of pairs of 511 keV photons detected in coincidence by detectors surrounding the patient. An accurate reconstruction requires in addition information on the spatial distribution of the attenuation coefficient for the 511 keV photons (the *attenuation image*), which is needed to compensate for the absorption or scattering of the photons by the tissues. Failure to correct for attenuation prevents accurate quantification of the tracer uptake and can also affect the qualitative interpretation of the images. This happens for instance when the increased tracer uptake in a localized lesion is almost exactly compensated by the increased attenuation caused by this lesion [1]. This paper deals with attenuation correction in time-of-flight (TOF) PET.

Various methods have been implemented to measure the attenuation image [2]. The state-of-the-art method uses a hybrid PET/CT scanner and obtains the attenuation image as a by-product of the diagnostic CT scan. The CT scan measures the attenuation coefficient at an average photon energy between 50 keV and 100 keV, these data are then extrapolated to the required energy of 511 keV to yield the attenuation image used for PET [3]. A benefit of this method is the good spatial resolution and signal-to-noise ratio (SNR) of the CT measurements compared to the PET emission data. However there are situations where the CT information is incomplete or inaccurate. In clinical practice, the most common source of bias is the geometrical mismatch between the emission data and the attenuation image, caused by patient motion or by different respiratory patterns in the CT and PET scans. This mismatch can be reduced by synchronizing the PET and CT scans with the respiratory motion (*gated scans*) but this increases the CT radiation exposure unless ultra-low dose protocols are applied [4]. Additional difficulties with CT-based attenuation correction arise when using radiological contrast agents [5], or when the axial field-of-view covered by the CT scan is limited to reduce the radiation dose [6].

An alternative solution for attenuation correction recently emerged with the hybrid PET/MR scanners. These scanners also provide anatomical images with good resolution and SNR but avoid radiation exposure and allow simultaneous acquisition of the PET and MR data. However, the relationship between the physical parameters measured in MR and the 511 keV attenuation coefficient is complex, and despite recent progresses [7–11] estimating the attenuation from MR data remains more prone to errors than with PET-CT scanners.

While the CT or MR data do not contain direct information about the activity image, the opposite is not true: the PET emission data do contain information about the attenuation image. This motivated the development of iterative algorithms, which simultaneously estimate the tracer distribution and the attenuation image from the combined emission and transmission data measured with external rotating positron sources [12–14]. A more ambitious approach is to dispose altogether with transmission data by estimating the attenuation image using only the emission data. Various algorithms have

been proposed to that effect [15–21], but they meet limited success unless strong prior knowledge is available, such as the assumption that the attenuation is uniform, which is an acceptable approximation for some applications such as brain imaging. This limited success is not surprising since it is known [22] that the solution to the simultaneous estimation is not unique in the absence of prior information.

The measurement of the time-of-flight in modern PET scanners increases the amount of information by a factor roughly equal to the ratio between the patient diameter and the width of the TOF profile, which is of the order of 8 cm in current scanners [23–25]. This additional information can be exploited to compress the data by reducing the number of angular [26, 27] or TOF [28] samples, and was also shown to improve the robustness of image reconstruction to errors in the attenuation image [29, 30]. Successful simultaneous reconstructions of attenuation and emission from combined PET and MR data [31–33] have also demonstrated the useful informative content of TOF-PET data.

This paper explores attenuation correction in TOF-PET using only the emission data. The study of an analytical model with continuous sampling and noise free data has shown [34] that the solution of the simultaneous estimation is unique in the sense that

- the activity image is uniquely determined up to a global multiplicative constant and,
- the *attenuation factors* (the exponential of minus the line integral of the attenuation image) are uniquely determined for all lines of response (LORs) which have activity, up to the reciprocal of the same multiplicative constant.

We consider here a discrete model of the problem and apply maximum likelihood (ML) estimation, which is expected to yield better results than an analytical model, owing to the high noise level in typical PET data. Two discrete ML approaches are possible.

- The first one maximizes the data likelihood with respect to the activity and attenuation images. This can be done by alternatively updating the activity and the attenuation images using the MLAA algorithm originally introduced for non-TOF PET [18]. The analysis of this algorithm for TOF-PET and its evaluation with simulated and measured data [35] opens promising perspectives for clinical applications. A similar approach is proposed in [36].
- The second ML approach maximizes the data likelihood with respect to the activity image and the *attenuation sinogram*, which is the set of attenuation factors for all LORs. Maximization can be achieved with the MLACF algorithm presented at the 2012 IEEE Medical Imaging Conference [37].

We investigate here this second approach. Both MLAA and MLACF aim at maximizing the data likelihood, but the two approaches are based on different parameterizations of the unknown quantities and are therefore not equivalent. A first difference is that the number of parameters to be estimated can be very different in 3D TOF-PET where the number of LORs in the attenuation sinogram is much larger than the number of voxels in the attenuation image. A second difference is the impossibility with MLACF to add regularizing penalties based on prior knowledge of the attenuation coefficients in some classes of tissues [38].

This is because MLACF only estimates attenuation factors, the attenuation image is only calculated after reconstruction as an optional byproduct required to estimate the scatter background or the geometric alignment with CT data.

This paper presents the derivation of the MLACF algorithm for the case where the background due to scatter and randoms events is negligible or has been pre-corrected. As will be shown in section 2 this assumption allows a closed form optimization with respect to the attenuation factors and leads to a simple iterative algorithm, which only involves updating the activity image and is similar to the standard ML-EM algorithm. Sections 2 and 3 present a mathematical and a numerical study of the convergence of the algorithm. The extension to the case with background is analysed in [39], the algorithm involves in that case alternately updating the attenuation sinogram and the activity image.

2. The MLACF algorithm

2.1. The joint likelihood for simultaneous estimation of the activity image and attenuation sinogram

Consider a scanner which histograms the coincidence data in N LORs, with T time-of-flight bins for each LOR. We denote the measured data as $\mathbf{y} = \{y_{i,t} \in I\!N_0, i = 1, \ldots, N, t = 1, \ldots, T\}$, where $y_{i,t}$ is the number of events detected for LOR i and time bin t, and $I\!N_0$ is the set of non-negative natural numbers. The activity image is parameterized as a vector $\lambda = \{\lambda_j \geq 0, j = 1, \ldots, M\}$. Any discrete parameterization can be used but one typically takes M voxel basis functions, and λ_j is then the tracer concentration in voxel j. The goal of PET is to estimate λ from the data $\mathbf{y}$.

The expectation value of the non-attenuated data is denoted $p_{i,t}$ and is related to the activity image by

$$p_{i,t} = \sum_{j=1}^{M} c_{i,j,t}\lambda_j \quad i = 1, \ldots, N \;\; t = 1, \ldots, T \tag{1}$$

where the system matrix elements $c_{i,j,t} \geq 0$ model the physics of the scanner. We assume throughout the paper that there is no background due to scatter or random coincidences, or that the data $\mathbf{y}$ have been pre-corrected but are nevertheless modeled, to first approximation, as independent Poisson variables. Quantities summed over the TOF index are denoted by omitting the t index, in particular we define $y_i = \sum_{t=1}^{T} y_{i,t}$, $c_{i,j} = \sum_{t=1}^{T} c_{i,j,t}$ and $p_i = \sum_{t=1}^{T} p_{i,t}$.

The attenuation factors, denoted a_i, are independent of the TOF bin. This is the key property that ultimately allows solving the simultaneous estimation. The attenuation factors are related to the attenuation image μ by

$$a_i = \exp\{-\sum_{j=1}^{M'} l_{i,j}\mu_j\} \quad i = 1, \ldots, N \tag{2}$$

where the system matrix elements $l_{i,j} \geq 0$ correspond to the parameterization used for μ, which is not necessarily the same as for λ. Equation (2) and the constraint $\mu_j \geq 0$ imply the natural constraint $0 < a_i \leq 1$ on the attenuation factors.

In this paper we seek to estimate λ and a from the emission data $\mathbf{y}$, and equation (2) will not be used. With this approach the attenuation factors are seen as nuisance parameters of the estimation problem since the objective of PET is the visualization of the tracer distribution, while CT or MR data provide anatomical details and are used for localizing the activity. The attenuation image may nevertheless be needed for some applications, such as the geometric alignment of the anatomical data with PET. Although μ can in principle be reconstructed from the attenuation factors a_i, such a reconstruction may be difficult in practice because the a_i can be estimated only for LORs containing activity. This issue will not be discussed here.

The expectation value of the measured attenuated data is $< y_{i,t} >= a_i p_{i,t}$ and the logarithm of the likelihood is

$$L(y, \lambda, a) = \sum_{i=1}^{N} \sum_{t=1}^{T} \{-a_i \, p_{i,t} + y_{i,t} \log(a_i \, p_{i,t})\} \tag{3}$$

with $p_{i,t}$ given by equation (1) and terms independent of λ and a have been omitted. The goal of ML estimation is to calculate a maximizer of the likelihood,

$$(\lambda^*, a^*) \in \arg \max_{0<a_i \leq 1, \lambda_j \geq 0} L(y, \lambda, a) \tag{4}$$

The notation $\in$ indicates that the maximizer is not unique. First there is the scale invariance $L(y, \lambda, a) = L(y, \alpha\lambda, a/\alpha)$ for any $\alpha > 0$, which has an obvious physical interpretation and also holds for the continuous model [34]. In addition, local maxima cannot be excluded because the log-likelihood (3) is not jointly concave in (a, λ) even though it is concave in a and in λ separately‡. The possibility of local maxima of the likelihood will be investigated numerically in section 3.

In the next section, the likelihood will be maximized with respect to a at fixed activity image λ, allowing to eliminate the attenuation factors from the set of parameters that must be estimated.

2.2. Maximizing the likelihood at fixed λ

Consider a fixed activity image λ. The log-likelihood (3) is the sum of N functions depending each on a single attenuation factor a_i, with first and second derivatives given by

$$\frac{\partial L(y, \lambda, a)}{\partial a_i} = \sum_{t=1}^{T} \left\{ -p_{i,t} + \frac{1}{a_i} y_{i,t} \right\} \quad , \quad \frac{\partial^2 L(y, \lambda, a)}{\partial a_i^2} = \sum_{t=1}^{T} \frac{-1}{a_i^2} y_{i,t} \leq 0. \tag{5}$$

For each LOR with activity, i.e. such that $y_i = \sum_t y_{i,t} > 0$, the log-likelihood has a unique maximizer

$$a_i^* = \frac{\sum_{t=1}^{T} y_{i,t}}{\sum_{t=1}^{T} p_{i,t}} = \frac{y_i}{p_i} \tag{6}$$

If in addition we impose the constraint $a_i \leq 1$, the maximizer becomes

$$\begin{aligned} a_i^* &= \frac{y_i}{p_i} \quad \text{if} \quad \frac{y_i}{p_i} \leq 1 \\ &= 1 \qquad \text{otherwise,} \end{aligned} \tag{7}$$

‡ the same holds when the likelihood is parametrized using λ and μ as with the MLAA approach.

as can be verified by noting that $\partial L(y,\lambda,a)/\partial a_i|_{a_i^*=1} \geq 0$ at the upper edge of the allowed domain $a_i \leq 1$ when $y_i \geq p_i$.

In this paper, we ignore the constraint $a_i \leq 1$ and always use $a_i^* = y_i/p_i$. The following lemma shows that this can be done provided $L(y,\lambda,a)$ has an unconstrained optimizer.

Lemma 1. Let

$$(a_c,\lambda_c) \in \arg \max_{0<a_i\leq 1,\ 0\leq\lambda_j} L(y,\lambda,a) \tag{8}$$

be a global maximizer of the likelihood with the constraint $a_i \leq 1$. If there exists a global maximizer without that constraint,

$$(a_u,\lambda_u) \in \arg \max_{0<a_i,\ 0\leq\lambda_j} L(y,\lambda,a) \tag{9}$$

then for some constant $K > 0$, $(a_u/K, K\lambda_u)$ is also a global constrained maximizer.

Proof. Define the maximum attenuation factor $K = \max(a_u) = \max_{i\in 1..N} y_i/p_i < \infty$. Adding a constraint (in this case $a_i \leq 1$) can never increase the value of the maximum, and therefore

$$L(y,\lambda_c,a_c) \leq L(y,\lambda_u,a_u) = L(y,\lambda^*,a^*) \tag{10}$$

where we define $a^* = a_u/K, \lambda^* = K\lambda_u$ and the last equality follows from the scale invariance. Thus $L(y,\lambda^*,a^*) \geq L(y,\lambda_c,a_c)$ and by construction $a_i^* \leq 1$, so (a^*,λ^*) is also a global constrained maximizer of the likelihood. □

When applicable Lemma 1 means that maximum likelihood solutions can be obtained by ignoring the constraint $a_i \leq 1$ and by scaling the solution a posteriori. The method derived in this paper is not applicable if $L(y,\lambda,a)$ has no global constrained optimizer and it is unclear whether such a situation is possible.

2.3. The reduced log-likelihood function $\tilde{L}$

Inserting the optimized attenuation factor (6) in the log-likelihood (3), and keeping only the terms that depend on λ, we are left with the problem of maximizing

$$L(y,\lambda,a^*) = \tilde{L}(y,\lambda) + \text{ terms independent of } \lambda \tag{11}$$

where the *reduced log-likelihood* is defined as

$$\tilde{L}(y,\lambda) = \sum_{i=1}^{N}\left\{-y_i\ \log(p_i) + \sum_{t=1}^{T} y_{i,t}\log(p_{i,t})\right\} = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t}\ \log\frac{p_{i,t}}{p_i} \tag{12}$$

We assume in sections 2.3 and 2.4 that $p_{i,t} > 0$ unless $y_{i,t} = 0$. When satisfied this condition implies also that $p_i > 0$ unless $y_i = 0$. The definition of $\tilde{L}(y,\lambda)$ when $p_{i,t} = 0$ or $p_i = 0$ will be discussed separately in section 2.5.

The gradient of the *reduced log-likelihood* (12) is

$$\frac{\partial\tilde{L}(y,\lambda)}{\partial\lambda_j} = \sum_{i=1}^{N}\left\{-\frac{y_i c_{i,j}}{p_i} + \sum_{t=1}^{T}\frac{y_{i,t}c_{i,j,t}}{p_{i,t}}\right\} \quad j = 1,\ldots,M \tag{13}$$

and the $M \times M$ Hessian matrix is

$$H_{j,k} = \frac{\partial^2 \tilde{L}(y,\lambda)}{\partial\lambda_j\,\partial\lambda_k} = \sum_{i=1}^{N}\left\{\frac{y_i\,c_{i,j}\,c_{i,k}}{p_i^2} - \sum_{t=1}^{T}\frac{y_{i,t}\,c_{i,j,t}\,c_{i,k,t}}{p_{i,t}^2}\right\} \quad j,k = 1,\ldots,M \quad (14)$$

The Hessian is not necessarily non-positive definite and one cannot exclude the existence of local maxima and of saddle points (beyond the undetermined global factor due to the scale invariance of the likelihood). However we recall the following result [37]. The proof is repeated for completeness.

Proposition 2. If the data are consistent the reduced log-likelihood has no local maximum other than the global maxima.

Proof. Consistent data (belonging to the range of the TOF-PET transform) can be written as

$$y_{i,t} = a_i^\dagger\, p_{i,t}^\dagger \text{ with } p_{i,t}^\dagger = \sum_{j=1}^{M} c_{i,j,t}\lambda_j^\dagger \quad i = 1,\ldots,N \;\; t = 1,\ldots,T \quad (15)$$

and $p_i^\dagger = \sum_t p_{i,t}^\dagger$, for some non-negative $\lambda_j^\dagger \geq 0$ and $a_i^\dagger > 0$. From (13) one sees that $\lambda^\dagger$ is a stationary point of the reduced log-likelihood and one verifies by inserting (15) in (12) that

$$\tilde{L}(y,\lambda^\dagger) = \sum_{i=1}^{N}\left\{-y_i\log y_i + \sum_{t=1}^{T} y_{i,t}\log y_{i,t}\right\} \quad (16)$$

In addition $\lambda^\dagger$ is a global maximizer. This can be seen by considering an arbitrary non-negative λ and by subtracting (12) from (16),

$$\tilde{L}(y,\lambda^\dagger) - \tilde{L}(y,\lambda) = \sum_{i=1}^{N}\left\{y_i\log\frac{p_i}{y_i} - \sum_{t=1}^{T} y_{i,t}\log\left(\frac{p_{i,t}}{y_{i,t}}\right)\right\} \quad (17)$$

By the concavity of the log function

$$\sum_t \frac{y_{i,t}}{y_i}\log\left(\frac{p_{i,t}}{y_{i,t}}\right) \leq \log\left(\sum_t \frac{y_{i,t}}{y_i}\frac{p_{i,t}}{y_{i,t}}\right) = \log\frac{p_i}{y_i} \quad (18)$$

and using this inequality in (17) shows that $\tilde{L}(y,\lambda^\dagger) \geq \tilde{L}(y,\lambda)$.

Consider now an arbitrary non-negative λ. Note first that $\lambda^t \cdot \nabla\tilde{L}(y,\lambda) = 0$ and calculate the scalar product

$$\begin{aligned} Z = (\lambda^\dagger)^t \cdot \nabla\tilde{L}(y,\lambda) &= \sum_{i=1}^{N}\left\{-\frac{y_i p_i^\dagger}{p_i} + \sum_{t=1}^{T}\frac{y_{i,t}p_{i,t}^\dagger}{p_{i,t}}\right\} \\ &= \sum_{i=1}^{N}\frac{1}{a_i^\dagger}\left\{-\frac{y_i^2}{p_i} + \sum_{t=1}^{T}\frac{y_{i,t}^2}{p_{i,t}}\right\} \geq 0 \end{aligned} \quad (19)$$

where we used the data consistency $a^\dagger\, p^\dagger = y$ (equation (15)) and the inequality follows from the convexity of the function x^2 because

$$\sum_{t=1}^{T}\frac{y_{i,t}^2}{p_{i,t}} = p_i\sum_{t=1}^{T}\frac{p_{i,t}}{p_i}(\frac{y_{i,t}}{p_{i,t}})^2 \geq p_i\left(\sum_{t=1}^{T}\frac{p_{i,t}}{p_i}\frac{y_{i,t}}{p_{i,t}}\right)^2 = \frac{y_i^2}{p_i} \quad (20)$$

Note that both in (18) and (19) the equality holds if and only if for all lines of response $i = 1, \ldots, N$ the ratios $p_{i,t}/y_{i,t}$ are independent of t. In that case $Z = 0$ and λ is also a global maximum of the reduced likelihood, i.e. $\tilde{L}(y, \lambda) = \tilde{L}(y, \lambda^\dagger)$.

The proof of Proposition 2 proceeds by contradiction. Suppose $\tilde{\lambda}$ is a local maximum of the reduced log-likelihood such that $\tilde{L}(y, \tilde{\lambda}) < \tilde{L}(y, \lambda^\dagger)$. Consider the restriction of $\tilde{L}$ to the segment linking $\lambda^\dagger$ and $\tilde{\lambda}$:

$$U(\alpha) = \tilde{L}(y, \lambda_\alpha = \alpha\lambda^\dagger + (1-\alpha)\tilde{\lambda}) \qquad 0 \leq \alpha \leq 1 \tag{21}$$

Since we assumed that $U(\alpha)$ has a local maximum in $\alpha = 0$ and is maximum in $\alpha = 1$, its derivative

$$U'(\alpha) = (\lambda^\dagger - \tilde{\lambda})^t \cdot \nabla \tilde{L}(y, \lambda_\alpha) \tag{22}$$

must take strictly negative values somewhere in the interval $(0, 1)$. But this is in contradiction with inequality (19) because for $0 < \alpha < 1$, the derivative can be rewritten as

$$\begin{aligned} U'(\alpha) &= \frac{1}{1-\alpha}\left((\lambda^\dagger)^t \cdot \nabla \tilde{L}(y, \lambda_\alpha) - (\lambda_\alpha)^t \cdot \nabla \tilde{L}(y, \lambda_\alpha)\right) \\ &= \frac{1}{1-\alpha}(\lambda^\dagger)^t \cdot \nabla \tilde{L}(y, \lambda_\alpha) \geq 0 \end{aligned} \tag{23}$$

and this is non-negative by (19). Therefore $\tilde{L}(y, \lambda) = \tilde{L}(y, \lambda^\dagger)$ and one concludes that with consistent data a local maximum is necessarily a global maximum. □

2.4. A surrogate approach and the MLACF algorithm

We consider some activity image $\tilde{\lambda}$ and the corresponding unattenuated data expectations

$$\tilde{p}_{i,t} = \sum_{j=1}^{M} c_{i,j,t}\tilde{\lambda}_j \quad i = 1, \ldots, N \;\; t = 1, \ldots, T \tag{24}$$

and $\tilde{p}_i = \sum_t \tilde{p}_{i,t}$. We assume in this section that all images are strictly positive, $\lambda_j > 0, j = 1, \ldots, M$. As will be seen in section 2.5 (Lemma 3) this assumption holds for the iterated solutions defined by the algorithm derived below.

Aiming at the iterative optimization of the reduced likelihood by optimization transfer [40], we build a function $\tilde{L}^{sur}(y, \lambda, \tilde{\lambda})$, with the usual surrogate properties

$$\begin{aligned} &\tilde{L}^{sur}(y, \tilde{\lambda}, \tilde{\lambda}) = \tilde{L}(y, \tilde{\lambda}) \\ &\tilde{L}^{sur}(y, \lambda, \tilde{\lambda}) \leq \tilde{L}(y, \lambda). \end{aligned} \tag{25}$$

Such a surrogate function can be obtained by considering separately the two terms in (12). For the *first term* the rooftop theorem for the concave log function yields

$$\log p_i \leq \log \tilde{p}_i + (p_i - \tilde{p}_i)\left(\frac{d \log p_i}{dp_i}\right)_{p_i = \tilde{p}_i} = \log \tilde{p}_i + \frac{p_i - \tilde{p}_i}{\tilde{p}_i} \tag{26}$$

and therefore

$$-y_i \log(p_i) \geq -y_i \log \tilde{p}_i - \frac{y_i\,(p_i - \tilde{p}_i)}{\tilde{p}_i} \tag{27}$$

For the *second term* in (12) define the quantities

$$w_{i,j,t} = \frac{c_{i,j,t}\tilde{\lambda}_j}{\tilde{p}_{i,t}} \qquad z_{i,j,t} = \frac{\lambda_j \tilde{p}_{i,t}}{\tilde{\lambda}_j} \tag{28}$$

which satisfy $w_{i,j,t} \geq 0$, $\sum_j w_{i,j,t} = 1$, and $\sum_j w_{i,j,t}\, z_{i,j,t} = p_{i,t}$. Since the log is concave, one has then for each i, t that

$$\log p_{i,t} = \log \left(\sum_j w_{i,j,t}\, z_{i,j,t} \right) \geq \sum_{j=1}^{M} w_{i,j,t}\, \log z_{i,j,t} \tag{29}$$

Putting the two terms together and summing over the data bins yields

$$\tilde{L}^{sur}(y, \lambda, \tilde{\lambda}) = \sum_{i=1}^{N} \left\{ -y_i\, \log \tilde{p}_i - \frac{y_i\, (p_i - \tilde{p}_i)}{\tilde{p}_i} + \sum_{t=1}^{T} \sum_{j=1}^{M} \frac{y_{i,t}\, c_{i,j,t} \tilde{\lambda}_j}{\tilde{p}_{i,t}} \log(\frac{\lambda_j \tilde{p}_{i,t}}{\tilde{\lambda}_j}) \right\} \tag{30}$$

One easily checks using (27) and (29) that this function satisfies the two properties (25) and is therefore a valid surrogate for the reduced likelihood. In addition $\tilde{L}^{sur}$ is separable (i.e. equal to a sum of terms, each of which depends on a single voxel λ_j) and hence easy to maximize.

The gradient of the surrogate is

$$\nabla \tilde{L}^{sur}_j = \frac{\partial \tilde{L}^{sur}(y, \lambda, \tilde{\lambda})}{\partial \lambda_j} = \sum_{i=1}^{N} \left\{ -\frac{y_i\, c_{i,j}}{\tilde{p}_i} + \sum_{t=1}^{T} \frac{y_{i,t}\, c_{i,j,t} \tilde{\lambda}_j}{\tilde{p}_{i,t}} \frac{1}{\lambda_j} \right\} \tag{31}$$

The Hessian matrix is diagonal and negative semi-definite,

$$\frac{\partial^2 \tilde{L}^{sur}(y, \lambda, \tilde{\lambda})}{\partial \lambda_j^2} = -\frac{\tilde{\lambda}_j}{\lambda_j^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{i,t}\, c_{i,j,t}}{\tilde{p}_{i,t}} \leq 0 \tag{32}$$

hence the surrogate is concave and has no local maxima.

Solving $\nabla \tilde{L}^{sur} = 0$ we obtain an iterative update which maps the estimate $\tilde{\lambda}$ on a new estimate λ [37]:

$$\lambda_j = \mathcal{T}(\tilde{\lambda})_j = \frac{\tilde{\lambda}_j}{\sum_{i=1}^{N} \frac{y_i\, c_{i,j}}{\tilde{p}_i}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{i,t} c_{i,j,t}}{\tilde{p}_{i,t}} \tag{33}$$

Applying this mapping iteratively with $\tilde{\lambda} = \lambda^k$ and $\lambda = \lambda^{k+1}$, where k denotes the iteration number, defines the MLACF algorithm. Some remarks are in order:

- The algorithm is closely related to the usual maximum-likelihood-expectation-maximization (ML-EM) algorithm for emission tomography with known attenuation. The similarity is best appreciated by rewriting (33) in the same form as ML-EM with attenuation factors $a_i^* = y_i/\tilde{p}_i$ defined according to (7):

$$\mathcal{T}(\tilde{\lambda})_j = \frac{\tilde{\lambda}_j}{\sum_{i,t} a_i^* c_{i,j,t}} \sum_{i,t} \frac{y_{i,t} c_{i,j,t}}{\tilde{p}_{i,t}} \tag{34}$$

- The curvature of the surrogate is high along axes j for which λ_j is small, resulting as with ML-EM in a slow convergence of the "cold" voxels.

- The algorithm is scale invariant: multiplying the initial estimate by a factor $\alpha > 0$ produces the same sequence of iterates, all multiplied by α. This is expected since the solution is determined only up to a factor. In practice, the activity image λ can be rescaled at each iteration, e.g. to guarantee that $\max_i a_i^* = 1$ or that $||\lambda|| = 1$. We chose the latter option and define the *normalized MLACF algorithm* with the mapping:

$$\mathcal{T}_{\mathcal{N}}(\tilde{\lambda}) = \frac{1}{||\mathcal{T}(\tilde{\lambda})||}\mathcal{T}(\tilde{\lambda}) \tag{35}$$

  with the L_2 norm $||x|| = (\sum_j x_j^2)^{1/2}$.
- In the non-TOF case, there is only a single time bin, $T = 1$. The mapping (33) reduces then to the $M \times M$ identity $\mathcal{T} = I\!I$.

2.5. Active data bins and the handling of zeroes

In the previous sections, the reduced log-likelihood (12), the surrogate (30), and the MLACF iteration (33) have been defined assuming that $p_{i,t} > 0$ and $p_i > 0$, unless the corresponding data are also zero.

Define the set of active data bins for a given voxel as:

$$\tau_j = \{(i,t) \,|\, c_{i,j,t}\, y_{i,t} > 0\} \quad j = 1, \ldots, M \tag{36}$$

and adopt an equivalent definition for the TOF-summed data,

$$\iota_j = \{i \,|\, c_{i,j}\, y_i > 0\} \quad j = 1, \ldots, M \tag{37}$$

Note that $\iota_j = \emptyset \Rightarrow \tau_j = \emptyset$. If for some voxel $\tau_j = \emptyset$, then

$$\frac{\partial \tilde{L}(y,\lambda)}{\partial \lambda_j} = \sum_{i \in \iota_j} \left\{ -\frac{y_i c_{i,j}}{p_i} \right\} \leq 0 \tag{38}$$

and therefore any maximizer of the reduced log-likelihood satisfies $\lambda_j = 0$ (because of the non-negativity constraint). If in addition $\iota_j = \emptyset$ the reduced log-likelihood does not depend on the value of voxel j, the maximizer is undefined and we chose the logical (but arbitrary) estimate $\lambda_j = 0$. These inactive voxels with $\tau_j = \emptyset$ can be set to zero and need not be further considered when maximizing $\tilde{L}(y,\lambda)$. We therefore assume from now on that $\tau_j \neq \emptyset$ for $j = 1, \ldots, M$. We will also assume that $y_i \geq 1$ and $y_{i,t} \geq 1$ for all non-zero data bins.

Having thus discarded the inactive voxels, one easily shows that the MLACF iterates are strictly positive:

Lemma 3. If $\lambda_j^0 > 0, j = 1, \ldots, M$, the MLACF iterates defined by the sequence $\lambda^{n+1} = \mathcal{T}_{\mathcal{N}}(\lambda^n)$ are strictly positive, i.e. $\lambda_j^n > 0, j = 1, \ldots, M$.

Proof. If $\lambda_j^n > 0, j = 1, \ldots, M$, then $p_{i,t}^n > 0$ for all (i,t), and hence also $p_i^n = \sum p_{i,t}^n > 0$. Therefore the MLACF update $\lambda^{n+1} = \mathcal{T}(\lambda^n)$ defined by (33) produces a positive image, $(\mathcal{T}(\lambda^n))_j > 0, j = 1, \ldots, M$. Applying this recursively, and starting from a positive initial image estimate λ^0 proves the lemma. □

The positivity property of Lemma 3 does not exclude that the MLACF iterates might converge to a limit point having some zero voxels, $\lambda_j^n \to 0$. One needs therefore to extent the definition of the reduced likelihood. If for some data bin (i,t), $y_{i,t} > 0$ and $p_i > 0$ but $p_{i,t} = 0$, it is natural to define $\tilde{L}(y,\lambda) = -\infty$, which is the limit of (12) when $p_{i,t} \to 0$. If in addition $p_i = 0$, the ratio $p_{i,t}/p_i$ and the reduced log-likelihood are undetermined. In this latter case ($p_i = 0$ and $y_i > 0$ for some LOR i), it is impossible (even by rescaling) to satisfy the constraint that $a_i^* = y_i/p_i \leq 1$ (see equation (6)).

2.6. Convergence of MLACF

The following proposition gives the main properties of the MLACF algorithm. The proof is given in appendix.

Proposition 4. Consider the sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with a positive initial image $\lambda_j^0 > 0$.

- The sequence of iterates is asymptotically regular, $||\lambda^{n+1} - \lambda^n|| \to 0$ as $n \to \infty$, and the reduced likelihood is non-decreasing, $\tilde{L}(y,\lambda^{n+1}) \geq \tilde{L}(y,\lambda^n)$.
- The sequence has a limit point λ^*, and $\nabla_j \tilde{L}(y,\lambda^*) = 0$ for any voxel satisfying $\lambda_j^* > 0$.
- All limit points of the sequence $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ have the same value of the reduced log-likelihood.

This result provides only a partial understanding of the convergence of MLACF. The limit points cannot be minima because $\tilde{L}$ is non-decreasing, but they might be saddle points or local maxima because there is no guarantee that the Hessian (14) is concave. Proposition 4 guarantees convergence of MLACF only under the restrictive additional assumptions that a) the likelihood has a unique global maximum $\lambda^\dagger$ such that $\lambda_j^\dagger > 0$ for all active voxels ($\tau_j \neq \emptyset$), and b) all limit points λ^* of the sequence of iterates are such that $\lambda_j^* > 0$ for all active voxels.

3. Numerical results

3.1. Simulation parameters

We digitized a 2D thorax phantom on a $M = 64 \times 64$ image with pixel size 8.027 mm. Simulated TOF-PET data are generated by forward projecting this phantom with radial pixel size 8.027 mm, 64 angular samples on $[0,\pi)$, and $T = 8$ times bins with sampling $\Delta\tau = 64.0$ mm. The TOF profile was a gaussian with FWHM 80 mm. The aim of this study is not to assess the practical value of MLACF but to get insight into its convergence and uniqueness properties, hence this coarse discretization was chosen to allow performing a very large number of iterations with various initial estimates λ^0. The activity and attenuation images are shown in Figure 1, the phantom support is an ellipse with axes 300 mm and 470 mm and the minimum attenuation factor a_i was 0.015. A vial with activity 0.5, diameter 40 mm, and water attenuation was added outside the phantom and used to scale the

reconstructed activity image at the end of the reconstruction. Poisson noise was added to the data to generate three noisy data sets S1, S2 and S3 with respectively a total of 479705, 15990 and 3198 events, corresponding to respectively 300, 10, and 2 events in the maximum data bin $< y_{i,t} >$. A large number of data bins are equal to zero in data set S3, allowing to challenge the algorithm's behaviour at the edge of the admissible domain. When generating the noisy data, a rescaling is applied after adding noise to ensure that $\sum y_i$ is the same for the three data sets S1,S2 and S3, up to statistical fluctuations.

The MLACF iteration was run up to 10^5 iterations (without using data subsets and with a matched backprojector), starting with a uniform image estimate $\lambda_j^0 = 1$ and with a set of 30 random initial images generated as $\lambda_j^0 = 0.1 + 0.9R$ where R is a pseudo-random number with uniform distribution in $(0, 1)$. All calculations were done in double precision. Initially we replaced the ratios $y_{i,t}/p_{i,t}$ and y_i/p_i in the numerator and denominator of the MLACF update (33) by zero when $p_{i,t}$ (respectively p_i) was smaller than some small ϵ, however we found that this margin is not needed and it is not used in the results presented here.

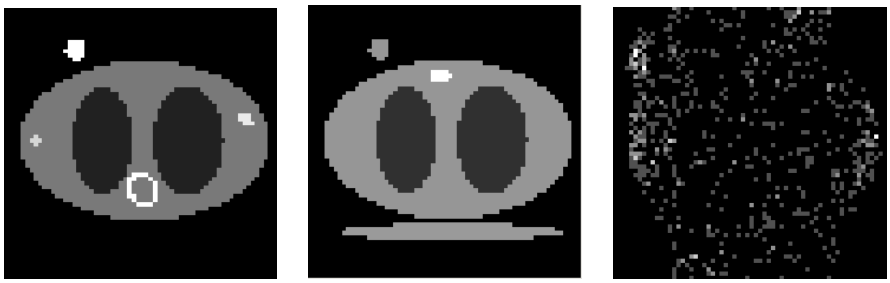

Figure 1. The simulated phantom. Emission (left): activity is 0.2 (background tissues), 1.7 ("heart"), 0.05 ("lungs"), 0.40 and 0.45 ("tumors"), and 0.5 (vial). Attenuation (middle): 0.00966/mm (background tissues and vial), 0.00266/mm ("lungs"), 0.0187/mm ("spine"), and 0.01/mm ("bed"). Right: sinogram of the central TOF bin of the noisiest data S3, illustrating the large number of bins (i, t) with zero counts. The displayed sinogram only has five different grey levels corresponding to 0,1,2,3 and 4 events.

3.2. Convergence

Figure 2 shows the activity image reconstructed starting from a uniform initial image. Figure 3, 4 and 5 show the convergence of the reduced log-likelihood for the three noise levels. We verified that $\tilde{L}(y, \lambda^n)$ monotonically increases for all data sets, initializations, and iterations, even though the convergence becomes slower with increasing noise level. Figure 3 and 5 also show the difference between the reduced log-likelihood obtained with a uniform estimate and with one particular random initialization.

In some cases (Figure 4), the convergence is irregular, with long sequences of iterates without significant improvement of the cost function. The second set of points in Figure 4 is obtained applying exactly the same algorithm but to a different realization S2' of the noisy data, for the same maximum count as S2. This result shows that the irregular

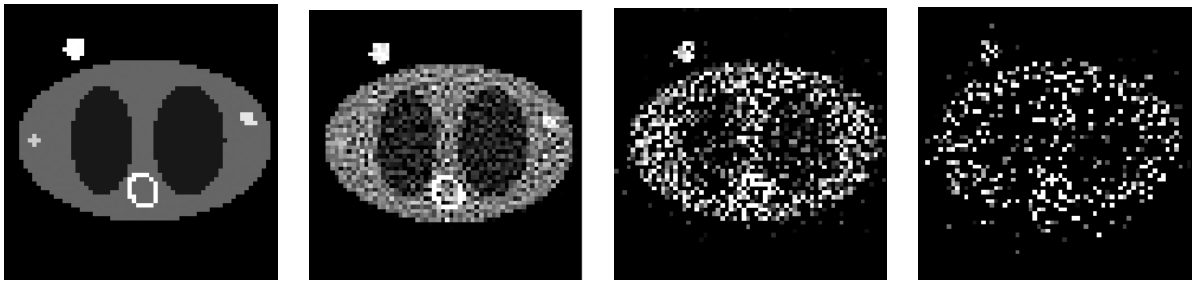

Figure 2. The activity image reconstructed from noise-free data and from the data sets S1, S2 and S3 (left to right). Grey scale $(0, 0.5)$.

convergence observed with S2 is specific to a particular noise realization. Figure 6 compares for that data set S2 the solution during that convergence plateau and at the last iteration. The convergence of the iterates shown in Figure 7 appears compatible with the asymptotic regularity of the algorithm predicted by Proposition 4.

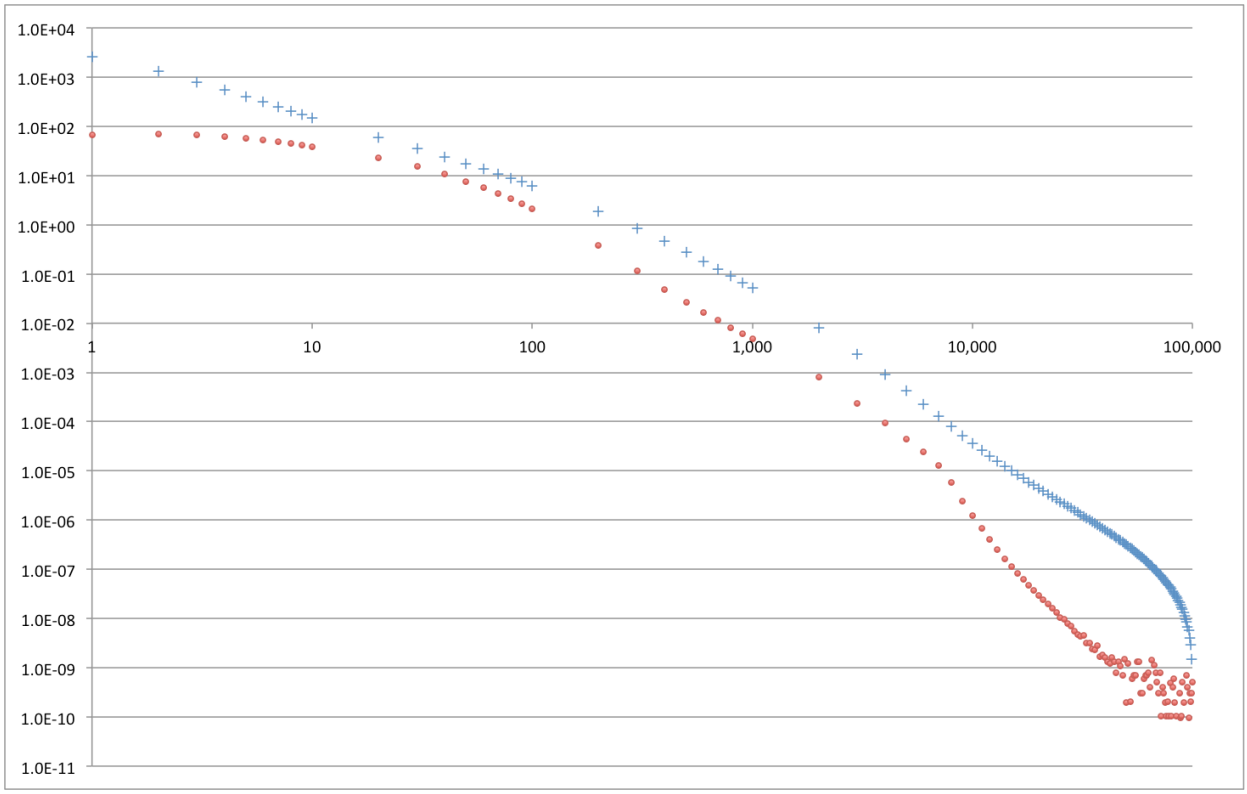

Figure 3. The reduced log-likelihood $-\tilde{L}(y, \lambda^n) + \tilde{L}(y, \lambda^{100000})$ for the low noise data set S1 (blue +), and the difference $|\tilde{L}(y, \lambda^n) - \tilde{L}(y, \lambda'^n)|$, (red circles), versus the number of iterations n. The sequences λ^n and λ'^n are obtained respectively with a uniform and one of the random initial images.

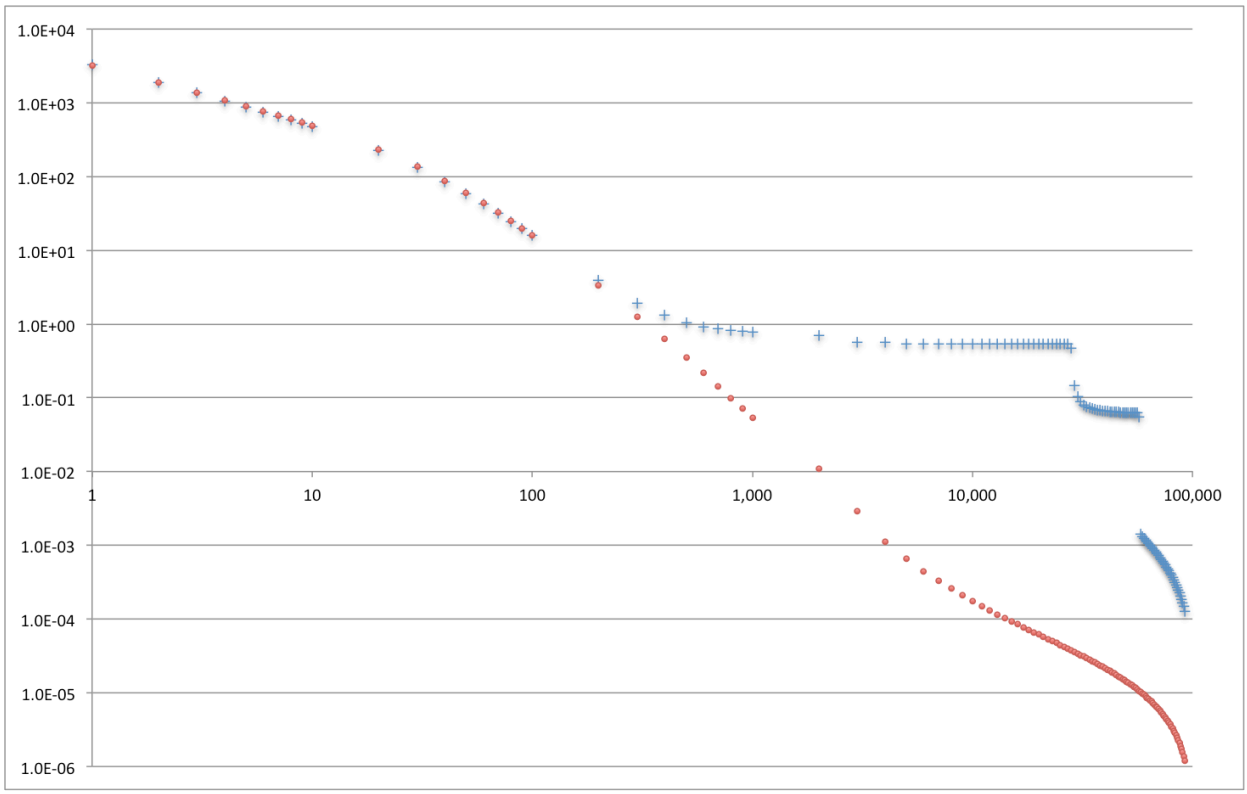

Figure 4. The reduced log-likelihood $-\tilde{L}(y,\lambda^n)+\tilde{L}(y,\lambda^{100000})$ for the medium noise data set S2 (blue +), and for another data set S2' with the same total count, (red circles), versus the number of iterations n. The sequences λ^n are obtained with a uniform initial images.

3.3. Uniqueness

Unless the data are consistent (see Proposition 2), we were not able to determine whether the reduced log-likelihood has a unique (up to the undetermined scale factor) global maximum. To get some insight into the possible existence of local maxima and of saddle points we compared the reconstructions obtained with the 31 initial activity images described in section 3.1. First we compare the final values of the reduced likelihood after 10^5 MLACF iterations. The ratio $(\max(\tilde{L}(y,\lambda^{10^5}))-\min(\tilde{L}(y,\lambda^{10^5}))/|\mathrm{mean}(\tilde{L}(y,\lambda^{10^5})|$ is equal to 5.7×10^{-14}, 4.2×10^{-9}, and 4.4×10^{-4} for the data sets S1, S2 and S3 respectively. Figure 8 shows a plot of the relative RMSE difference $||\lambda-\lambda'||/||\lambda||$ between all pairs of images λ,λ' reconstructed for each data set using different initializations, with the values sorted by increasing magnitudes. The RMSE differences are small for S1 and do not show any particular structure, and the same behaviour is observed for S2 (not shown). For the noisiest data set S3, however, the RMSE are larger and the plot shows some structure. This structure might suggest convergence to different images, but might also be caused by the slow convergence or by numerical errors. For very noisy data sets the question of uniqueness remains open, though Figure 9 shows that the differences between reconstructions obtained with different initial images are visually small.

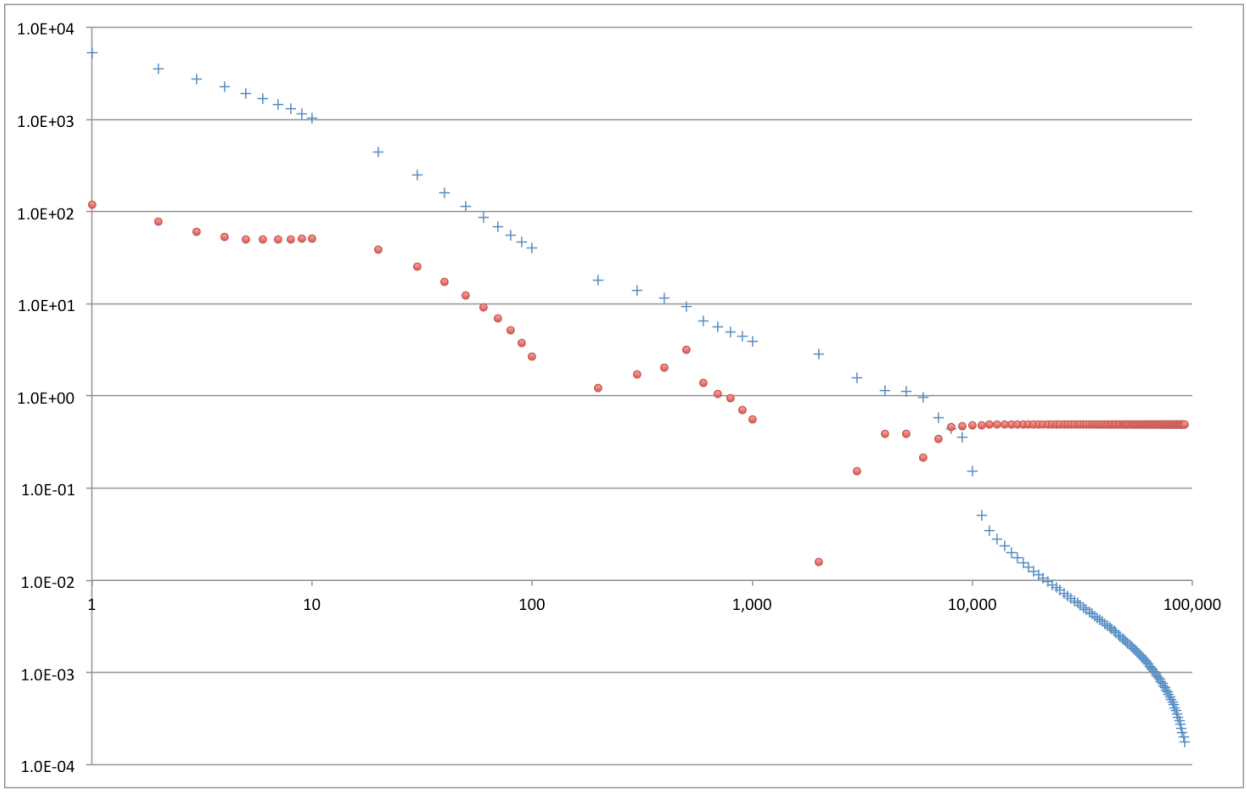

Figure 5. The reduced log-likelihood $-\tilde{L}(y,\lambda^n)+\tilde{L}(y,\lambda^{100000})$ for the high noise data set S3 (blue +), and the difference $|\tilde{L}(y,\lambda^n)-\tilde{L}(y,\lambda'^n)|$, (red circles), versus the number of iterations n. The sequences λ^n and λ'^n are obtained respectively with a uniform and one of the random initial images.

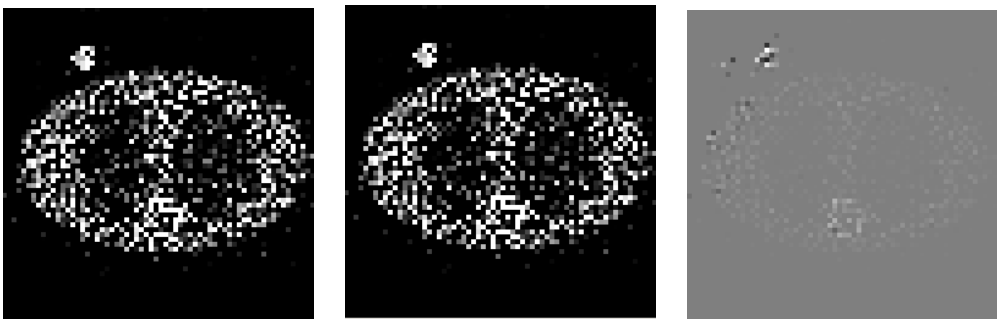

Figure 6. The activity image reconstructed from data set S2 with 10^5 (left) and 5000 (center) iterations, with grey scale $(0, 0.5)$. The right image is the difference, grey scale $(-0.015, +0.015)$.

3.4. Comparison with OSEM

The MLACF algorithm estimates the activity image without using any information on the attenuation. To evaluate the influence this lack of information has on convergence, we reconstructed the same data using ML-EM, assuming that the exact value of the attenuation

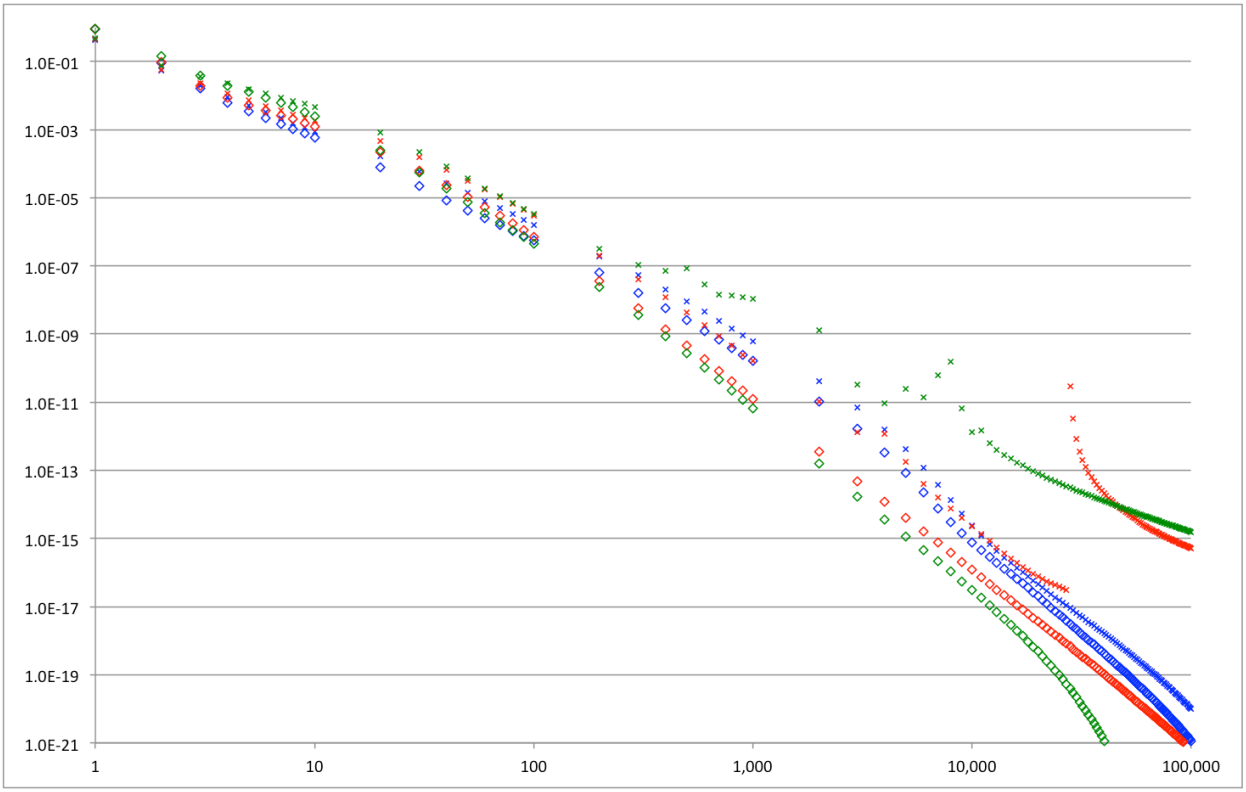

Figure 7. The value of $||\lambda^{n+1} - \lambda^n||^2/||\lambda^n||^2$ versus iteration number n for data set S1 (blue), S2 (red), and S3 (green). The sequences λ^n are obtained with the ML-EM algorithm with known attenuation factors a (diamonds) and with MLACF (x), both starting with a uniform initial image.

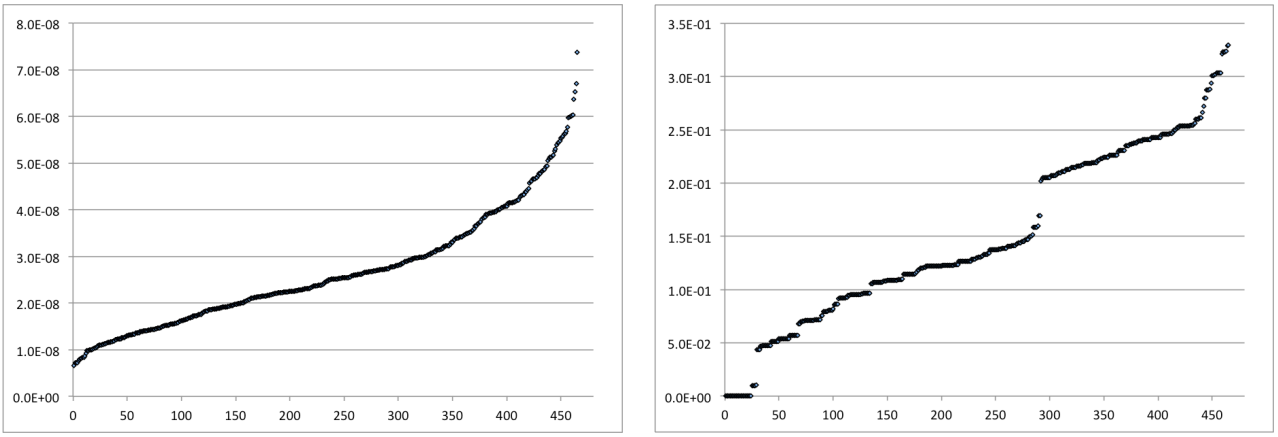

Figure 8. The relative RMSE difference between all pairs among 31 reconstructions from data set S1 (left) and among 31 reconstructions from data set S3 (right). The values are sorted by increasing magnitudes. All reconstructions use 10^5 MLACF iterations but starting with a uniform and with random initial images.

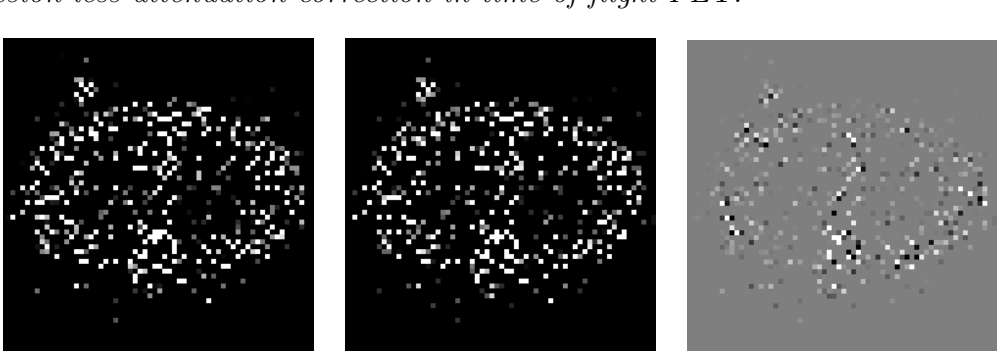

Figure 9. The activity image reconstructed from the data set S3 with the two random initial images that resulted in the largest RMSE difference (0.33). Grey scale $(0, 0.5)$. Right: Difference image, grey scale $(-0.5, +0.5)$.

Table 1. Relative RMSE differences between reconstructed and exact activity images (uniform initial image, 10^5 iterations).

	noise free	data S1 max. count 300	data S2 max. count 10	data S3 max. count 2
ML-EM vs. phantom	8.53×10^{-6}	2.48×10^{-1}	9.24×10^{-1}	1.66
MLACF vs. phantom	1.93×10^{-5}	2.05×10^{-1}	1.16	1.54
MLACF vs. ML-EM	1.64×10^{-5}	1.94×10^{-1}	6.42×10^{-1}	8.58×10^{-1}

factors a_i is known:

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{\sum_{i=1}^N a_i c_{i,j}} \sum_{i=1}^N \sum_{t=1}^T \frac{y_{i,t} c_{i,j,t}}{p_{i,t}^n} \tag{39}$$

The ML-EM implementation uses the same ingredients as MLACF (matched backprojector, double precision arithmetic). Figure 7 shows that ML-EM has a faster and more regular convergence, and the difference is more marked for the low count case. Since the goal of this work is to study convergence, the ML-EM and MLACF reconstructions have not been regularized. Nevertheless, Figure 10 shows that the reconstructed ML-EM and MLACF activity images have similar noise levels, though as expected the ML-EM reconstructions are better. This observation and the RMSE data in Table 1 suggest that the redundancy of the TOF data is sufficient to avoid dramatic image degradation by the absence of attenuation information.

Figure 11 compares the log-likelihood (3) obtained with ML-EM and MLACF. The values for MLACF are calculated as $L(y, \lambda, a^*(\lambda)) = \tilde{L}(y, \lambda) + \sum_i (-y_i + y_i \log(y_i))$. The likelihood converges to a larger value with MLACF than with ML-EM since MLACF does not fix the attenuation factors a priori and therefore has more degrees of freedom available to maximize the likelihood. For the same reason the difference between MLACF and ML-EM increases as the number of counts decreases.

Finally, the three data sets have been reconstructed using 100000 iterations of ML-EM, with the same 31 initial images as used in section 3.3. In contrast with the similar

comparison for MLACF in section 3.3, the relative RMSE differences between pairs of images reconstructed for each data set are all lower than 2×10^{-8}, even for the noisiest case S3. This observation is expected since ML-EM is known to converge to the unique maximizer of the likelihood, but it suggests that the RMSE shown in Figure 8 can probably not be attributed to numerical errors.

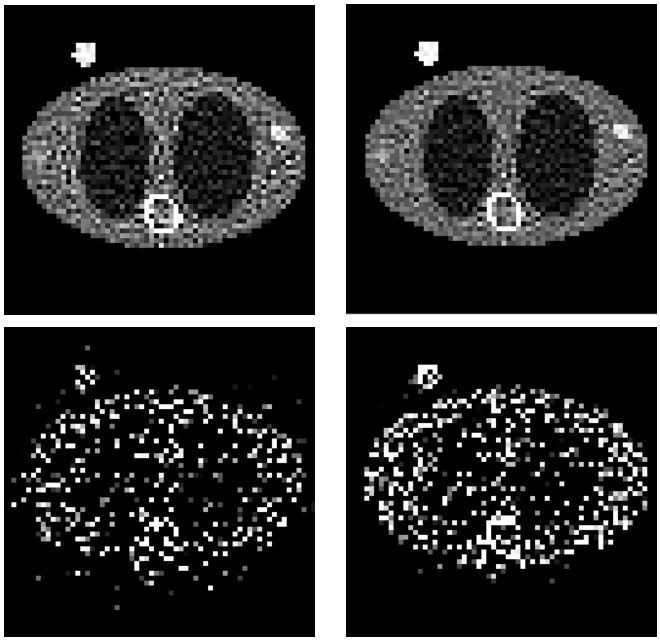

Figure 10. The activity image reconstructed from data set S1 (top row) and S3 (bottom row) with 10^5 iterations. Left: MLACF. Right: ML-EM with known attenuation. Grey scale $(0, 0.5)$.

4. Conclusion

Several groups have shown that time-of-flight PET allows to simultaneously estimate the activity and attenuation images. These results open new perspectives for various applications including for example stand-alone PET, PET-CT studies where patient motion prevents the direct utilization of the CT data for attenuation correction, PET-CT studies with a reduced CT field-of-view, and also potential applications in PET-MR.

The maximum likelihood method for this problem can use as unknown parameter either the attenuation image or the attenuation factors. Promising results have been obtained with the two approaches, but fundamental questions remain concerning the uniqueness of the maximum likelihood estimator and the convergence of the iterative algorithms. The difficulty with this non-linear inverse problem stems from the non-concavity of the likelihood, and we were only able to prove the absence of local maxima of the likelihood for the case where the data are consistent and the solution is parametrized using λ and a.

This paper dealt with the second approach, which takes the attenuation factors

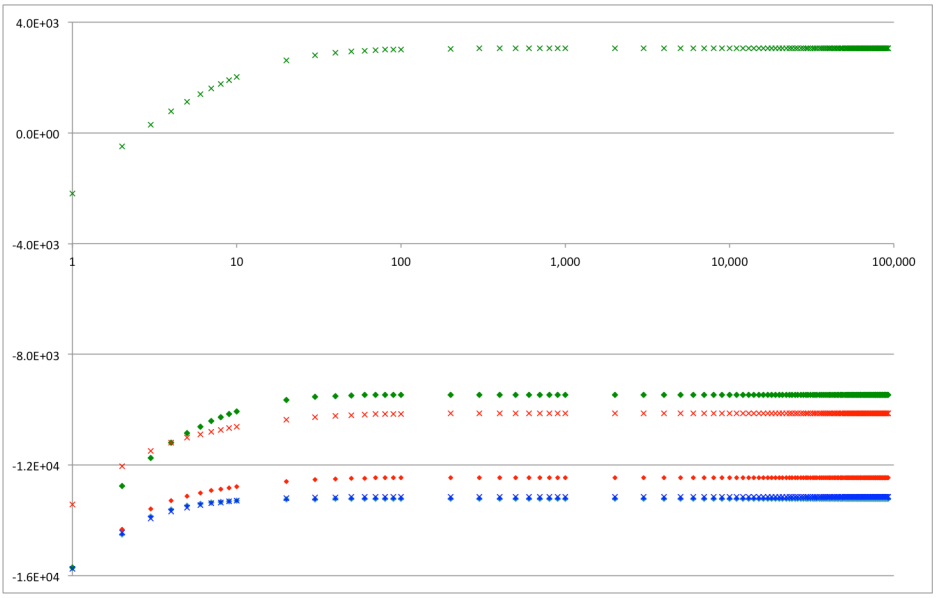

Figure 11. The log-likelihood $L(y, a, \lambda^n)$ versus iteration number n for data set S1 (blue), S2 (red) and S3 (green). The sequences λ^n are obtained with the ML-EM algorithm (diamonds) with known attenuation factors a, and with MLACF ($\times$), both starting with a uniform initial image.

and activity image as unknown parameters. We assumed that the scatter and random background is negligible and showed that this assumption allows eliminating the attenuation parameters. This led to an iterative algorithm, MLACF [37], which only updates the activity image. Attractive properties of MLACF include its simplicity and similarity with ML-EM, the monotonic increase of the likelihood, and the asymptotic regularity (the norm of the solution increments tends to zero). Convergence however could not be proven in general. We conducted simulation studies with the unregularized MLACF algorithm and with a number of iterations much larger than would be applied in practice. The results give reasonable confidence that MLACF converges in practice to a unique maximizer of the likelihood, except for extremely noisy data sets. Additional studies are warranted for that case.

Acknowledgements

This work is supported by a grant of the Fonds voor Wetenchappelijk Onderzoek Vlaanderen (F.W.O), by a KU Leuven research grant (GOA/11/006) and by Siemens Healthcare. The authors thank Girish Bal, Mike Casey, Maurizio Conti, Christian Michel, Vladimir Panin and Charles Watson (Siemens Healthcare, Molecular Imaging, Knoxville, TN) for useful discussions.

5. Appendix: convergence of the MLACF algorithm.

Let $\lambda^n, n = 0, 1, 2, \ldots$ be the sequence of iterates generated by a positive initial estimate $(\lambda^0)_j > 0$ and by the update step $\lambda^n = \mathcal{T}_{\mathcal{N}}(\lambda^{n-1})$ defined by the normalized mapping (35). This appendix presents the proof of proposition 4. The general line of the analysis follows the scheme in [41].

The following positive quantities will be used:

$$\begin{aligned} \xi_j &= \min_{i \in \iota_j} c_{i,j} > 0 \quad , \quad \xi = \min_j \xi_j > 0 \\ \eta_j &= \min_{(i,t) \in \tau_j} c_{i,j,t} > 0 \quad , \quad \eta = \min_j \eta_j > 0 \\ \omega &= \max_i \sum_j c_{i,j} \quad , \quad \sigma = \max_{i,t} \sum_j c_{i,j,t} > 0 \end{aligned} \tag{40}$$

Lemma 5. The sequence $\tilde{L}(y, \lambda^n), n = 0, 1, 2, \ldots$ is non-decreasing and converges.

Proof. The iterates are obtained by optimization transfer and the surrogate properties (25) guarantee by construction that the sequence $\tilde{L}(y, \lambda^n), n = 0, 1, 2, \ldots$ is non-decreasing. The positivity of the λ^n (Lemma 3) and of the system matrix elements $c_{i,j,t}$ guarantee that $p^n_{i,t} \leq p^n_i$, and therefore

$$\tilde{L}(y, \lambda^n) = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t} \log \frac{p^n_{i,t}}{p^n_i} \leq 0 \tag{41}$$

so that the sequence $\tilde{L}(y, \lambda^n)$ is bounded above by 0. A non-decreasing upper bounded sequence converges. □

Lemma 6. Let $\tilde{\lambda} \in \mathbb{R}^M$ be any positive vector. The following inequality holds for the non-normalized mapping $\mathcal{T}$ in (33):

$$\tilde{L}(y, \mathcal{T}(\tilde{\lambda})) - \tilde{L}(y, \tilde{\lambda}) \geq \frac{1}{2} \sum_j \left(\mathcal{T}(\tilde{\lambda}) - \tilde{\lambda} \right)^2_j \frac{\tilde{\lambda}_j}{\bar{\lambda}^2_j} \sum_{(i,t) \in \tau_j} \frac{y_{i,t} c_{i,j,t}}{\tilde{p}_{i,t}} \tag{42}$$

where $\bar{\lambda} = \alpha \mathcal{T}(\tilde{\lambda}) + (1 - \alpha)\tilde{\lambda}$ for some $0 \leq \alpha \leq 1$.

Proof. For any positive $x \in \mathbb{R}^M$ and a fixed $\tilde{\lambda}$, we can represent $\tilde{L}^{sur}(y, x, \tilde{\lambda})$ by a Taylor expansion around its maximizer $\lambda = \mathcal{T}(\tilde{\lambda})$. Using the fact that $\nabla_\lambda \tilde{L}^{sur}(y, \lambda, \tilde{\lambda}) = 0$,

$$\tilde{L}^{sur}(y, x, \tilde{\lambda}) = \tilde{L}^{sur}(y, \lambda, \tilde{\lambda}) + \frac{1}{2}(x - \lambda)^t \cdot \nabla^2 \tilde{L}^{sur}(y, \bar{\lambda}, \tilde{\lambda}) \cdot (x - \lambda) \tag{43}$$

where the Hessian is taken at some point $\bar{\lambda}$ on the segment linking λ to x. Rewriting this equation at $x = \tilde{\lambda}$, using from (25) the surrogate properties $\tilde{L}^{sur}(y, \tilde{\lambda}, \tilde{\lambda}) = \tilde{L}(y, \tilde{\lambda})$ (for the LHS) and $\tilde{L}^{sur}(y, \lambda, \tilde{\lambda}) \leq \tilde{L}(y, \lambda)$ (for the RHS), one obtains

$$\tilde{L}(y, \tilde{\lambda}) \leq \tilde{L}(y, \lambda) + \frac{1}{2}(\tilde{\lambda} - \lambda)^t \cdot \nabla^2 \tilde{L}^{sur}(y, \bar{\lambda}, \tilde{\lambda}) \cdot (\tilde{\lambda} - \lambda) \tag{44}$$

Using the diagonal non-positive Hessian (32) concludes the proof. □

Lemma 7. Let $\tilde{\lambda} \in \mathbb{R}^M$ be any positive vector with $||\tilde{\lambda}|| = 1$.

$$\tilde{L}(y, \mathcal{T}(\tilde{\lambda})) - \tilde{L}(y, \tilde{\lambda}) \geq C \frac{||\mathcal{T}(\tilde{\lambda}) - \tilde{\lambda}||^2}{\max(1, ||\mathcal{T}(\tilde{\lambda})||)} \tag{45}$$

where $\mathcal{T}$ is the mapping (33) and $C = (1/2)\min(\xi/\omega, \eta/\sigma) > 0$.

Proof. Define $\lambda = \mathcal{T}(\tilde{\lambda})$ and denote the curvature factor in (42) as

$$\rho_j = \frac{\tilde{\lambda}_j}{(\alpha\lambda_j + (1-\alpha)\tilde{\lambda}_j)^2} \sum_{(i,t)\in\tau_j} \frac{y_{i,t} c_{i,j,t}}{\tilde{p}_{i,t}} \tag{46}$$

Using the MLACF update (33) one has the equivalent expression

$$\rho_j = \frac{\lambda_j}{(\alpha\lambda_j + (1-\alpha)\tilde{\lambda}_j)^2} \sum_{i\in\iota_j} \frac{y_i c_{i,j}}{\tilde{p}_i} \tag{47}$$

To obtain a lower bound on ρ_j we consider separately the two cases $\lambda_j \geq \tilde{\lambda}_j$ and $\lambda_j < \tilde{\lambda}_j$.

- If $\lambda_j \geq \tilde{\lambda}_j$, then $\bar{\lambda}_j \leq \lambda_j$ because $0 \leq \alpha \leq 1$, and therefore, using (47),

$$\rho_j \geq \frac{1}{\lambda_j} \sum_{i\in\iota_j} \frac{y_i c_{i,j}}{\tilde{p}_i} \tag{48}$$

  Using $\tilde{p}_i = \sum_k c_{i,k}\tilde{\lambda}_k \leq \omega \max_k \tilde{\lambda}_k$, and noting that $\tilde{\lambda}_k \leq ||\tilde{\lambda}|| = 1$, and that $y_i \geq 1$ for $i \in \iota_j$, one obtains

$$\rho_j \geq \frac{1}{\lambda_j}\frac{\xi_j}{\omega} \geq \frac{1}{||\lambda||}\frac{\xi}{\omega} \tag{49}$$

  where we used $\lambda_j \leq ||\lambda||$.

- If $\lambda_j < \tilde{\lambda}_j$, then $\bar{\lambda}_j \leq \tilde{\lambda}_j$ and therefore, using (46),

$$\rho_j \geq \frac{1}{\tilde{\lambda}_j} \sum_{(i,t)\in\tau_j} \frac{y_{i,t} c_{i,j,t}}{\tilde{p}_{i,t}} \tag{50}$$

  Using $\tilde{p}_{i,t} = \sum_k c_{i,k,t}\tilde{\lambda}_k \leq \sigma \max_k \tilde{\lambda}_k$, and noting that $\tilde{\lambda}_k \leq ||\tilde{\lambda}|| = 1$, and that $y_{i,t} \geq 1$ for $(i,t) \in \tau_j$, one obtains

$$\rho_j \geq \frac{\eta_j}{\sigma} \geq \frac{\eta}{\sigma}. \tag{51}$$

Inserting the lower bounds (49) and (51) into (42) yields inequality (45). □.

Lemma 8. Asymptotic regularity. Starting with any positive initial estimate $\lambda_j^0 > 0$, the sequence of *normalized* iterates $\lambda^{n+1} = \mathcal{T}_{\mathcal{N}}(\lambda^n)$ is such that

$$\begin{aligned} &\lim_{n\to\infty} ||\mathcal{T}(\lambda^n) - \lambda^n|| = 0 \\ &\lim_{n\to\infty} ||\mathcal{T}(\lambda^n)|| = 1 \\ &\lim_{n\to\infty} ||\mathcal{T}_{\mathcal{N}}(\lambda^n) - \lambda^n|| = 0 \end{aligned} \tag{52}$$

Proof. Applying Lemma 7 with $\tilde{\lambda} = \lambda^n$ yields

$$\frac{||\mathcal{T}(\lambda^n) - \lambda^n||^2}{\max(1, ||\mathcal{T}(\lambda^n)||)} \leq e_n \tag{53}$$

with $e_n = (1/C)(\tilde{L}(y, \mathcal{T}(\lambda^n)) - \tilde{L}(y, \lambda^n))$. From Lemma 5, the sequence of reduced log-likelihoods converges (recall that the value of the reduced log-likelihood is not modified by the normalization), therefore $\lim_{n\to\infty} e_n = 0$. Defining $z_n = \mathcal{T}(\lambda^n) - \lambda^n$ inequality (53) becomes

$$||z_n||^2 \leq e_n \max(1, ||\lambda^n + z_n||) \leq e_n \max(1, 1 + ||z_n||) \leq e_n (1 + ||z_n||) \tag{54}$$

where we used $||\lambda^n|| = 1$ and the triangular inequality. Inequality (54) can be rewritten as

$$(||z_n|| - e_n/2)^2 \leq e_n + e_n^2/4 \tag{55}$$

and finally

$$||z_n|| \leq e_n/2 + \sqrt{e_n + e_n^2/4} \tag{56}$$

The upper bound in the RHS tends to zero because $e_n \to 0$, hence $||z_n|| \to 0$. This proves the first equation of the Lemma. The second equation of the Lemma follows because the iterates λ^n are normalized so that, again using the triangular inequality,

$$1 - ||\mathcal{T}(\lambda^n) - \lambda^n|| \leq ||\mathcal{T}(\lambda^n)|| \leq 1 + ||\mathcal{T}(\lambda^n) - \lambda^n|| \tag{57}$$

Finally,

$$\begin{aligned} ||\mathcal{T}_{\mathcal{N}}(\lambda^n) - \lambda^n|| &= ||\frac{\mathcal{T}(\lambda^n)}{||\mathcal{T}(\lambda^n)||} - \lambda^n|| \leq ||\frac{\mathcal{T}(\lambda^n)}{||\mathcal{T}(\lambda^n)||} - \mathcal{T}(\lambda^n)|| + ||\mathcal{T}(\lambda^n) - \lambda^n|| \\ &\leq |(1 - ||\mathcal{T}(\lambda^n)||)| + ||\mathcal{T}(\lambda^n) - \lambda^n|| \end{aligned} \tag{58}$$

We have already shown that the two terms in the RHS tend to zero, therefore the LHS tends to zero, which proves the last equation in (52). □

Lemma 9. If a positive sequence $\lambda^n \in \mathbb{R}^M$ converges to some limit point λ^* such that $\lambda_j^* > 0$, then

$$\begin{aligned} \lim_{n\to\infty} \frac{y_i\, c_{i,j}}{p_i^n} &= \frac{y_i\, c_{i,j}}{p_i^*} \\ \lim_{n\to\infty} \frac{y_{i,t}\, c_{i,j,t}}{p_{i,t}^n} &= \frac{y_{i,t}\, c_{i,j,t}}{p_{i,t}^*} \end{aligned} \tag{59}$$

with $p_i^n = \sum_k c_{i,k}\lambda_k^n$, $p_i^* = \sum_k c_{i,k}\lambda_k^*$, $p_{i,t}^n = \sum_k c_{i,k,t}\lambda_k^n$ and $p_{i,t}^* = \sum_k c_{i,k,t}\lambda_k^*$.

Proof. Take any LOR such that $y_i c_{i,j} > 0$. Consider the quantity

$$\Delta^n = \left| \frac{1}{\sum_k c_{i,k}\lambda_k^n} - \frac{1}{\sum_k c_{i,k}\lambda_k^*} \right| \leq \left| \frac{\sum_k c_{i,k}(\lambda_k^* - \lambda_k^n)}{(c_{i,j})^2 \lambda_j^n \lambda_j^*} \right| \tag{60}$$

Since $\lambda_j^n \to \lambda_j^*$, there is an integer N such that for each $n > N$, $\lambda_j^n > (1/2)\lambda_j^* > 0$. Therefore,

$$\Delta^n \leq \left| 2\frac{\sum_k c_{i,k}(\lambda_k^* - \lambda_k^n)}{(c_{i,j})^2 (\lambda_j^*)^2} \right| \tag{61}$$

The denominator is a positive number independent of n, and as $\lambda^n \to \lambda^*$, one concludes that $\Delta^n \to 0$. The first relation in (59) immediately follows. The proof of the second relation is similar.□

Proposition 4. Consider the sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with a positive initial image $\lambda_j^0 > 0$.

- The sequence of iterates is asymptotically regular, $||\lambda^{n+1} - \lambda^n|| \to 0$ as $n \to \infty$, and the reduced likelihood is non-decreasing, $\tilde{L}(y, \lambda^{n+1}) \geq \tilde{L}(y, \lambda^n)$,
- The sequence of iterates has an limit point λ^*, and $\nabla_j \tilde{L}(y, \lambda^*) = 0$ for any voxel satisfying $\lambda^*_j > 0$,
- All limit points of the sequence $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ have the same value of the reduced log-likelihood.

Proof. The asymptotic regularity and monotonicity are given by Lemmas 5 and 8. The λ^n are normalized, hence their sequence is bounded and contains a converging subsequence $\lambda^{n(s)}, s = 1, 2, \ldots$. Denote λ^* the limit of this subsequence.

We now consider some voxel j such that $\lambda^*_j > 0$ and prove that $\nabla_j \tilde{L}(y, \lambda^*) = 0$. Recall that $\lambda^{n(s)}_j > 0$ for all j by Lemma 3. Define

$$d^n_j = \sum_{i=1}^{N} \frac{y_i \, c_{i,j}}{p^n_i} \quad \text{and} \quad v^n_j = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{i,t} c_{i,j,t}}{p^n_{i,t}} \tag{62}$$

and using (33) note that

$$\lambda^{n+1}_j = (T_{\mathcal{N}}(\lambda^n))_j = \frac{1}{||T(\lambda^n)||} T(\lambda^n)_j = \frac{1}{||T(\lambda^n)||} \lambda^n_j \frac{v^n_j}{d^n_j} \tag{63}$$

Applying this and using (13),

$$\begin{aligned} \lambda^n_j \nabla_j \tilde{L}(y, \lambda^n) &= \lambda^n_j(-d^n_j + v^n_j) \\ &= (\lambda^{n+1}_j - \lambda^n_j) d^n_j + (1 - \frac{1}{||T(\lambda^n)||}) \lambda^n_j v^n_j \end{aligned} \tag{64}$$

On the other hand one has from (62) the identity

$$\sum_{j=1}^{M} \lambda^n_j d^n_j = \sum_{j=1}^{M} \lambda^n_j v^n_j = \sum_{i=1}^{N} y_i := Y. \tag{65}$$

Combined with the non-negativity of all quantities involved this implies that $\lambda^n_j d^n_j \leq Y$ and $\lambda^n_j v^n_j \leq Y$.

Consider a voxel such that $\lambda^*_j > 0$. Since $\lambda^{n(s)}_j \to \lambda^*_j$, there is an iteration number S such that for $s > S$, the subsequence is sufficiently close to convergence, so that $\lambda^{n(s)}_j > \lambda^*_j/2$. Equation (64) leads then, for $s > S$, to

$$\begin{aligned} (\lambda^*_j/2)|\nabla_j \tilde{L}(y, \lambda^{n(s)})| &\leq \lambda^{n(s)}_j \, |\nabla_j \tilde{L}(y, \lambda^{n(s)})| \\ &\leq |\lambda^{n(s)+1}_j - \lambda^{n(s)}_j| \frac{Y}{(\lambda^*_j/2)} + |1 - \frac{1}{||T(\lambda^{n(s)})||}| \, Y \end{aligned} \tag{66}$$

The two terms in the RHS tend to zero because by Lemma 8 the sequence λ^n is asymptotically regular and $||T(\lambda^n)|| \to 1$. Therefore $\nabla_j \tilde{L}(y, \lambda^{n(s)}) \to 0$. By Lemma 9, it follows that $\nabla_j \tilde{L}(y, \lambda^*) = \lim_s \nabla_j \tilde{L}(y, \lambda^{n(s)}) = 0$.

Finally, two limit points λ^* and $\lambda^\dagger$ of the bounded sequence λ^n have the same value of the reduced log-likelihood: this is an immediate consequence of Lemma 5. □

References

[1] Bai C, Kinahan P E , Brasse D , Comtat C, Townsend D W, Meltzer C C, Villemagne V, Charron M, and Defrise M 2003 Effects of attenuation on tumor detection in wholebody PET oncology imaging *Journal of Nuclear Medicine* **44** 1855-1861

[2] Meikle S R and Badawi R D 2005 Quantitative techniques in PET, in "Positron emission tomography, basic sciences", eds. Bailey D L, Townsend D W, Valk P E, Maisey M N, Springer 2005, 93-126.

[3] Kinahan P E, Townsend D W, Beyer T, and Sashin D 1998 Attenuation correction for a combined 3D PET/CT scanner *Med Phys* **25 (10)** 2046-2053.

[4] Xia T, Alessio A M, De Man B, Manjeshwar R, Asma E, Kinahan P E 2012 Ultra-low dose CT attenuation correction for PET/CT *Phys Med Biol* **57** 309328.

[5] Carney J, Beyer T, Brasse D, Yap J T, Townsend D W 2002 CT-Based Attenuation Correction for PET/CT Scanners in the Presence of Contrast Agent *Records 2002 IEEE Nuclear Science Symposium and Medical Imaging Conference, Norfolk.*

[6] Panin V, Defrise M, Nuyts J, Rezaei A, Casey M E 2012 Reconstruction of uniform sensitivity emission image with partially known axial attenuation information in PET-CT scanners, *Records 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim* 2166-2173.

[7] Schreibmann E, Nye JA, Schuster DM, Martin DR, Votaw J, Fox T 2010 MR-based attenuation correction for hybrid PET-MR brain imaging systems using deformable image registration *Med Phys* **37(5)** 2101-2109

[8] Keereman V, Fierens Y, Broux T, Deene Y D, Lonneux M, and Vandenberghe S 2010 MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences *J Nucl Med* **51 (5)** 812-818

[9] Hofmann M, Bezrukov I, Mantlik F, Aschoff P, Steinke F, Beyer T, Pichler BJ, Schlkopf B 2011 MRI-based attenuation correction for whole-body PET/MRI: quantitative evaluation of segmentation- and atlas-based methods *J Nucl Med* **52(9)** 1392-9

[10] Martinez-Moller A, Souvatzoglou M, Delso G, Bundschuh R A, Chefd'hotel C, Ziegler SI, Navab N, Schwaiger M, Nekolla S G 2009 Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI, *J Nucl Med* **50** 520-6

[11] Keereman V, Mollet P, Berker Y, Schulz V, Vandenberghe S 2012 Challenges and current methods for attenuation correction in PET/MR, *Mag Res Mat in Phys Biol and Med* **26** 81-98

[12] Clinthorne N H, Fessler J A, Hutchins G D and Rogers W L 1991 Joint maximum likelihood estimation of emission and attenuation densities in PET *Proc. IEEE Trans. Nucl. Sci. Symp. Med. Im. Conf.* **3** 1927-1932

[13] Erdogan H and Fessler J A 1999 Joint estimation of attenuation and emission images from PET scans *Records 1999 IEEE Nuclear Science Symposium and Medical Imaging Conference, Seattle* 1672-1675

[14] Glatting G, Wuchenauer M and Reske S N 2000 Simultaneous iterative reconstruction for emission and attenuation images in positron emission tomography, *Med. Phys* **27 (9)** 2065-2071

[15] Censor Y, Gustafson D E, Lent A and Tuy H 1979 New approach to the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients *IEEE Trans Nucl Sc* **27 (2)** 2775-2779.

[16] Natterer F and Herzog H 1992 Attenuation correction in positron emission tomography *Mathematical Methods in the Applied Sciences* **15** 321-330

[17] Welch A, Campbell C, Clackdoyle R, Natterer F, Hudson M, Bromiley A, Mikecz P, Chillcot F, Dodd M, Hopwood P, Craib S, Gullberg G T and Sharp P 1998 Attenuation correction in PET using consistency information, *IEEE Trans. Nuclear Science* **45 (6)** 3124-3141

[18] Nuyts J, Dupont P, Stroobants S, Benninck R, Mortelmans L, Suetens P 1999 Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms *IEEE Trans Med Imag* **18** 393-403

[19] Bronnikov A V 2000 Reconstruction of attenuation map using discrete consistency conditions *IEEE Trans Med Imaging* **19(5)** 451-462

[20] Krol A, Bowsher J E, Manglos S H, Feiglin D H, Tornai M P and Thomas FD 2001 An EM algorithm for estimating SPECT emission and transmission parameters from emission data only *IEEE Trans Med Imaging* **20** 218-232

[21] De Pierro A, Crepaldi F 2006 Activity and attenuation recovery from activity data only in emission computed tomography *Comput Appl Math* **25** 2-3

[22] Natterer F 1987 Attenuation correction in emission tomography *Inverse Problems* 21-23 ed. Sabatier, Academic Press

[23] Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS, 2007, Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities, *J Nucl Med.* **48(3)**, 471-80.

[24] Jacoby B W, Bercier Y, Conti M, Casey M E, Bendriem B, Townsend D W 2011 Physical and clinical performance of the mCT time-of-flight PET/CT scanner, *Phys. Med. Biol.*, **56**, pp. 2375-89.

[25] Bettinardi V, Presotto L, Rapisarda E, Picchio M, Gianolli L, Gilardi MC, 2011, Physical performance of the new hybrid PET-CT Discovery-690, *Med Phys.*, **38 (10)** 5394-411.

[26] S. Vandenberghe, M. E. Daube-Witherspoon, R. M. Lewitt and J. S. Karp, 2006, Fast reconstruction of 3D TOF PET data by axial rebinning and transverse mashing, *Phys. Med. Biol.*, **51**, 1603-1621.

[27] Defrise M, Panin V Y, Michel C, Casey M E 2008 Continuous and Discrete Data Rebinning in Time-of-Flight PET *IEEE Trans Med Imag* **27** 1310-1322.

[28] Cho S, Ahn S, Li Q and Leahy R 2009, Exact and approximate Fourier rebinning of PET data from time-of-flight to non time-of-flight, *Phys Med Biol* **54** 467-484.

[29] Turkington T G, Wilson J M 2009 Attenuation artifacts and time-of-flight PET, *Record Nuclear Science Symposium Conference (NSS/MIC)* 2297-2299.

[30] Conti M 2011 Why is TOF PET reconstruction a more robust method in the presence of inconsistent data? *Phys Med Biol* **56** 155-168

[31] Salomon A, Schulz V, Brinks R, Schweizer B, Goedicke A and Aach T 2009 Iterative generation of attenuation maps in TOF-PET/MR using consistency conditions *J Nucl Med* **50 (2)** 2013

[32] Salomon A, Goedicke A, Schweizer B, Aach T, Schulz V 2011 Simultaneous Reconstruction of Activity and Attenuation for PET/MR *IEEE Trans Med Imag* **30** 804-813

[33] Li H, El Fakhri G, Li Q 2013 Direct estimation of attenuation sinogram using TOF PET data and anatomical information, *Records 12th International Meeting on Fully 3D Reconstruction in Radiology and Nuclear Medicine, Lake Tahoe (CA)* 404-407

[34] M. Defrise, A. Rezaei, and J. Nuyts, 2012, Time-of-flight PET data determine the attenuation sinogram up to a constant, *Phys Med Biol* **57 (4)** 885-899.

[35] A. Rezaei, Defrise M, Bal G, Michel C, Conti M, Watson C, Nuyts J 2012 Simultaneous reconstruction of activity and attenuation in time-of-flight PET", *IEEE Trans Med Imaging* **31 (12)** 2224-2233.

[36] Ahn S, Qian H, Manjeshwar R M, 2012, Convergent iterative algorithms for joint reconstruction of activity and attenuation from time-of-flight PET data, *Records 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim*, 3695-3700.

[37] Nuyts J, Rezaei A, Defrise M, 2012, ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors, *Records 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim*, 2147-49.

[38] Nuyts J, Dupont P, Stroobants S, Maes A, Mortelmans L, Suetens P 1999 Evaluation of maximum-likelihood based attenuation correction in positron emission tomography *IEEE Trans Nucl Sci* **46 (4)** 1136-1141

[39] Rezaei A, Defrise M, Nuyts J, 2013, ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors, submitted to *IEEE Transactions on Medical Imaging.*

[40] Lange K, Hunter D R, Yang I 2000 Optimization transfer algorithms using surrogate objective functions. *J. Comp. Graph. Stat.* **9** 1-59

[41] De Pierro A R 1995 A Modified Expectation Maximization. Algorithm for Penalized Likelihood *IEEE Trans Med Imaging* **14 (1)** 132-7